DOCUMENT RESUME

ED 478 491 TM 035 091

AUTHOR	Bassiri, Dina; Schulz, E. Matthew
TITLE	Constructing a Universal Scale of High School Course Difficulty. ACT Research Report Series.
INSTITUTION	American Coll. Testing Program, Iowa City, IA.
REPORT NO	ACT-RR-2003-04
PUB DATE	2003-06-00
NOTE	34p.; For a related paper, see ED 455 248.
AVAILABLE FROM	ACT Research Report Series, P.O. Box 168, Iowa City, IA 52243-0168.
PUB TYPE	Reports - Research (143)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	College Entrance Examinations; Grade Point Average; *Grades (Scholastic); Grading; *High School Students; High Schools; Item Response Theory; *Prediction; *Rating Scales; Scaling
IDENTIFIERS	*ACT Assessment; *Rasch Model

ABSTRACT

This study examined the usefulness of applying the Rasch rating scale model (D. Andrich, 1978) to high school grade data. ACT assessment test scores (English, Mathematics, Reading, and Science) were used as "common items" to adjust for different grading standards in individual high school courses both within and across schools. This scaling approach yielded an ACT Assessment-adjusted high school grade point average (AA-HSGPA) that was comparable across schools, cohorts, and among students within the same school and cohort who take different courses. The AA-HSGPA was constructed for all ACT-tested students (N=36,652) in 50 selected high schools. First-year college grades at a large public university were available for approximately 1,500 of these students. AA-HSGPA was a better predictor of first-year college grade point average (GPA) than the regular high school grade point average (HSGPA). As expected, the regression of CGPA on HSGPA for high schools grouped by difficulty with regard to grading policy (easy or hard) differed, but the regressions of CGPA on AA-HSGPA and the ACT Composite score (ACTA) did not. The best model for predicting CGPA included both the ACT Composite score and AA-HSGPA. (Contains 4 tables, 5 figures, and 22 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ACT Research Report Series 2003-4

ED 478 491

Constructing a Universal Scale of High School Course Difficulty

Dina Bassiri

E. Matthew Schulz

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

P. Farrant

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM035091

BEST COPY AVAILABLE

ACT

2

June 2003

For additional copies write:
ACT Research Report Series
P.O. Box 168
Iowa City, Iowa 52243-0168

© 2003 by ACT, Inc. All rights reserved.

Constructing a Universal Scale of High School Course Difficulty

Dina Bassiri
E. Matthew Schulz

Abstract

This study examined the usefulness of applying the Rasch rating scale model (Andrich, 1978) to high school grade data. ACT Assessment test scores (English, Mathematics, Reading, and Science) were used as "common items" to adjust for different grading standards in individual high school courses both within and across schools. This scaling approach yielded an ACT Assessment-adjusted high school grade point average (AA-HSGPA) that was comparable across schools, cohorts, and among students within the same school and cohort who take different courses. The AA-HSGPA was constructed for all ACT-tested students (N=36,652) in 50 selected high schools. First-year college grades at a large public university were available for approximately 1,500 of these students. AA-HSGPA was a better predictor of first-year college grade point average (CGPA) than the regular high school grade point average (HSGPA). As expected, the regression of CGPA on HSGPA for high schools grouped by difficulty with regard to grading policy (easy or hard) differed, but the regressions of CGPA on AA-HSGPA and the ACT Composite score (ACTC) did not. The best model for predicting CGPA included both the ACT Composite score and AA-HSGPA.

Constructing a Universal Scale of High School Course Difficulty

As a measure of academic achievement, the grade point average (GPA) is limited by the extraneous effects of schools and courses. The problem of school effects was first to be recognized. Linn (1966) noted that variation in grading practices among high schools made the high school GPA (HSGPA) a sub-optimal predictor of achievement in college. The problem of course effects has received attention more recently. Students can earn a higher grade point average simply by taking easier courses. These problems fuel grade inflation at the high school level (Ziomek and Svec, 1995), and create incentives for students to avoid courses in difficult subjects such as mathematics and science (Johnson, 1997).

In an effort to improve the prediction of college GPA (CGPA) by HSGPA, a number of earlier studies attempted to adjust both the predictor and criterion grades (Bashaw, 1965; Bloom and Peters, 1961; Lindquist, 1963; Potthoff, 1964; Tucker, 1963). This approach, called "central prediction systems," was an attempt to control for school effects. Disappointment with this approach centered largely on the practical difficulties of its implementation. It was difficult to find enough students from a given high school within a given college to estimate coefficients in nested regression equations. Data for one or more variables in the equations were frequently missing for many of the students who could be found. Linn (1966) noted that the accuracy of central prediction systems often was not so much better than that of simply using a standardized test, such as the ACT Assessment, in conjunction with HSGPA, to predict college achievement as to justify the cost and expense of these systems.

In an attempt to construct more reliable, predictable measures of college achievement, scaling models have been used to control for the effects of college courses (Johnson, 1997; Young, 1990). These models have been applied only to college course grade data, possibly

because course-taking patterns differ more among college students than among high school students. Scale-adjusted CGPA measures, like CGPA, represent student achievement on a single (uni-dimensional) scale and are constructed entirely from course grade data.

These scaling procedures have shown promise. Compared to raw CGPA, scale-adjusted CGPA is more predictable by pre-admissions variables such as HSGPA and standardized test scores (Caulkins, Larkey, & Wei, 1996; Johnson, 1997; Young, 1990;). Scale-adjusted CGPA is also better able to predict which of two students received the higher grade if both students took the same college course (Johnson, 1997).

The question to be addressed by the present study is whether a scaling approach might be effective in controlling for both school and course effects on HSGPA. With regard to controlling for course effects *within* a high school, the motivation and procedures are similar to those in studies with college course grade data. For example, within any given high school, some college-bound students take core college-preparatory courses, while others don't. These differences reduce the ability of HSGPA to predict college achievement. A scale-adjusted HSGPA can be constructed from data within a high school because courses taken by the same students are "common items", and students who take the same courses are "common persons" in scaling terminology.

In order to control for school effects in a scaling procedure, however, one needs one or more common items between the schools. Linn (1966) suggested that an external variable, such as a standardized test, could be used in this fashion. In this study, we will use ACT Assessment test scores as "common items." The ACT Assessment tests (English, Mathematics, Reading, and Science) are curriculum-based achievement tests (ACT, 1997). ACT test scores have strong relationships to high school course loads and to self-reported grades in high school (Noble &

McNabb, 1989). Scores on these tests are comparable across high schools and over time (ACT, 1997). The ACT Assessment is taken by over one million high school juniors and seniors in thousands of high schools (ACT, 2002). With these characteristics, ACT Assessment tests are ideal "common items" for constructing a universal scale of high school course difficulty. The scale-adjusted HSGPA in our study is called "AA-HSGPA" because it is adjusted for school effects by the ACT Assessment tests.

A potential advantage of a scaling approach over a central prediction system concerns the sample size and data requirements for estimating prediction equations. An AA-HSGPA can be constructed for any student in a high school in which a sufficient number of students take the ACT Assessment. With some scaling models, as few as thirty students may be needed to estimate the difficulty of a course on the universal scale. These data can be collected over time, including more than one year, as long as the grading practices for the courses or school remain approximately the same. The AA-HSGPA can be used to predict achievement at *any* college. Within a college, AA-HSGPA can be used to predict achievement of students from different high schools. There does not have to be a minimum number of students from a given high school who enrolled in a particular college to incorporate school effects in CGPA predictions.

Source of Data

ACT Assessment scores include four subject-area scores—English, Mathematics, Reading, and Science—on a 1 to 36 point scale, plus a Composite score. The Composite score is the rounded average of the four subject area scores. The mean Composite score among ACT-tested students nationally is 20.8 (ACT, 2002).

Self-reported high school grades came from the Course Grade Information Section (CGIS) of the Student Profile component of the ACT Assessment. This component is completed

voluntarily by students when they register for the ACT Assessment. Students are asked to report grades in standard college preparatory courses such as English 9 to English 12, Algebra, Geometry, etc. Course grades are reported on the A to F scale. Previous research shows that self-reported grades are highly comparable to actual grades received (Sawyer, Laing & Houston, 1988). For all analyses high school course grades were numerically coded as A=4, B=3, C=2, D=1, and F=0. All analyses used self-reported grades in the 23 college preparatory high school courses included in the CGIS whose contents are represented by ACT Assessment tests (Table 1). Self-reported grades on non-core courses such as art and language were not included.

Data for this study included different samples, as explained below. Figures 1 and 2 depict the relationship among the different samples.

Sample A

This sample consisted of all of the high schools represented in four freshman cohorts from a public university (referred to as Midwest University or MU). There were 390,179 students from over 1,300 high schools. These students were in the high school graduating classes of 1996, 1997, 1998, or 1999, had self-reported high school grades and ACT Assessment scores, and were attending a high school that had one or more graduates who attended the Midwest University between 1997 and 2000 inclusively. In short, Sample A represents ACT-tested students in high schools that send students to Midwest University.

Sample B

Fifty schools were selected that were representative of the distribution of 'difficulty' of the 1300 plus high schools and had ten or more graduates who attended the Midwest University between 1997 and 2000 inclusively. The selection of these fifty schools was done in the following manner. The mean ACT Composite score (ACTC) and mean high school GPA, based

on self-reported grades, was computed for each of the 1300-plus schools (for 390,179 students). Mean HSGPA was regressed on mean ACTC. Fifty schools were selected so the weighted mean and variance of the residuals from the 50 schools matched the weighted mean and variance of the residuals from the 1300 plus schools. These residuals were weighted by the number of students within each school who attended Midwest University. This sample included all ACT-tested students who graduated from the fifty selected high schools from 1996 to 1999. There were 36,652 students in this sample.

Sample C

This sample included all Sample B students who earned at least one grade at the Midwest University in the two semesters of the year following their high school graduation. There were 1,505 of these students. The number of students per school (50 schools) in this sample ranged from 8 to 131 (for any given school, Sample C is a subset of Sample B and may contain fewer than 10 students).

Computation of College GPA (CGPA). The college GPA was the unweighted average of the grades reported for the first two semesters. Grades included pluses (+) and minuses (–). These were numerically coded as A=4, B=3, C=2, D=1, F=0 with 0.33 added or subtracted for a plus or minus (there were no F- grades). A CGPA was computed for all students in Sample C. The number of students for whom a CGPA was computed, by cohort, was 348 for 1996, 370 for 1997, 397 for 1998, and 390 for 1999. The number of grades per student ranged from 2 to 18. Over eighty-five percent of these students had 7, 8, or 9 grades. The mean CGPA, by cohort, was 2.72 for 1996 and 1997, 2.74 for 1998, and 2.80 for 1999.

Sample D (Students from Easy High Schools) and Sample E (Students from Hard High Schools)

Two or more schools were selected from each end of the ordered residuals from the regression of mean ACT Composite on mean HSGPA to represent easy and hard high schools with regard to grading policy of the fifty high schools. Easy and hard high schools were selected in an attempt to show that HSGPA is influenced by school effects, whereas, ACTC and AA-HSGPA are not. Note that negative residuals correspond to schools with strict grading policies (hard high schools), whereas positive residuals correspond to schools with lenient grading policies (easy high schools). Out of the 1,505 students included in this study who attended Midwest University, easy and hard high schools represented a total of 166 and 169 students, respectively.

Methods

Scaling Analyses

All scaling analyses were performed with the Bigsteps computer program, Version 2.27 (Wright & Linacre, 1990). Bigsteps estimates the parameters of a number of Rasch models including the Rating Scale Model (RSM) (Andrich, 1978) and the Partial Credit Model (PCM) (Masters, 1982). In a comparison study that involved a cross-validity component (Lei, Bassiri, & Schulz, 2003), the RSM and PCM performed well on course grade data in comparison to the other models including linear models (Caulkins, et al. 1996), the Graded Response Model (Samajima, 1969), and the Generalized Partial Credit Model (Muraki, 1992).

For J+1 distinct grades numerically coded F=0, D=1, …, A=4, (J+1 = 5) the RSM is:

$$\ln\left(\frac{P_{nij}}{P_{nij-1}}\right) = \beta_n - \delta_i - \tau_j, \qquad j=1,2,\ldots,J;\ i=1,\ldots,I;\ n=1,\ldots,N \qquad (1)$$

where

ln is the natural log function,

P_{nij} is the probability that student n earns grade j in course i,

P_{nij-1} is the probability that student n earns grade j-1 in course i,

β_n is the AA-HSGPA of student n,

δ_i is the difficulty of course i,

τ_j is the "step threshold"--the difficulty of earning grade j or higher given the student earned at least grade j-1.

In the RSM, courses share a common set of step thresholds (τ_j, j=0,...,J). Course-specific step thresholds are incorporated into the PCM by the use of a parameter (δ_{ij}) in place of δ_i-τ_i in Equation (1). In both models, the parameter value corresponding to j=0 is 0. The RSM was used for course grade data because the number of students per course within a high school was too small to yield reliable course-specific step threshold estimates.

To construct AA-HSGPA measures, we performed three basic series of Bigsteps analyses. Each series of analyses used Sample B--all ACT-tested students from 50 selected high schools. In these analyses, we recoded ACT test scores, which normally range from 1 to 36, into a smaller number of ordered categories, similar to course grade data. The ACT scale scores were re-coded as follows: [0,16]=0, [17,19]=1, [20,22]=2, [23,25]=3, [26,36]=4. This coding yielded the same number of levels as course grades (five), with approximately equal numbers of students per level. It was reasoned that the uniform relative frequency of the levels would maximize the collective reliability of the threshold parameter estimates and that more or fewer than five levels would give, respectively, too much or too little weight to the ACT tests in determining the characteristics of the common scale. Moderate departures from this coding scheme, such as the

use of four or six levels, or coding that produce lower frequencies at the extremes, were not expected to substantially affect the results of the study.

In the analyses described below, the PCM was used for the recoded ACT test scores because, as common items, sample sizes for these items were large enough to estimate item-specific step threshold parameters:

Step 1. Calibration of ACT tests. The purpose of this analysis was to bring the ACT tests onto a scale that represents high school course grades. This was done by jointly calibrating (estimating the parameters of) ACT tests and high school courses. A single Bigsteps analysis was performed using all the Sample B data. We did not consider it necessary at this point for the high school course parameters to be school-specific, so only twenty-seven items were defined for the analysis. The Partial Credit model was used for the 4 ACT tests and the Rating Scale model was used for the 23 high school courses. The overall estimate of the difficulty of each course provided by this analysis was used for purposes of general comparison (see Table 1), but not for any other purpose.

Step 2. Calibration of courses. School-specific analyses were performed in order to allow courses to have school-specific locations on the common scale. In each analysis, parameters of the four ACT test "items" were anchored (fixed) to the estimates obtained for them in Step 1, while Rating Scale model parameters of the 23 courses (and step parameters) were estimated. These analyses treated the ACT tests as common items across schools and treated courses with the same title, such as Algebra, in different schools as distinct courses. Each school had its own Rating Scale model step structure. Differences in course parameters reflect main effects of schools, courses, and their interaction. We refer to these effects loosely as between-school and within-school course effects.

Step 3. Estimating student measures. The person statistics output from the Step 2 analyses were not used because they were based on a combination of ACT test scores and high school grades. We preferred to work with measures that separated these two sources of information. In order to obtain a measure on the common scale based exclusively on course grades (except for the use of the ACT tests to create the common scale), analyses were performed again separately by school. Each analysis involved just the twenty-three courses. The course parameters were anchored at the estimates obtained in Step 2. The person measures obtained from these analyses (i.e., the $\hat{\beta}_n$) were AA-HSGPA measures. It should be noted that these measures are on a logit (log-odds) scale whose origin (zero) is the average item difficulty from the Step 1 analysis, and have a theoretical range of minus to plus infinity.

Regression Analyses

Regression analyses were performed on data from Samples C, D (easy) and E (hard) as follows:

Midwest University prediction analyses. In order to compare and contrast the ability of HSGPA and AA-HSGPA with or without ACTC in predicting CGPA, five prediction analyses (using proc REG in SAS) were performed on Sample C data.

1) CGPA = ACTC
2) CGPA = HSGPA
3) CGPA = ACTC, HSGPA
4) CGPA = AA-HSGPA
5) CGPA = ACTC, AA-HSGPA

Dummy variable regression. In an attempt to show that HSGPA is influenced by school effects, whereas, ACTC and AA-HSGPA are not, nine regression analyses (using proc GLM in

SAS) were performed on the Samples D (easy high schools) and E (hard high schools), where high school type (HS-type), easy or hard, was dummy coded as [hard =1] and [easy=0]. For the ACTC the regression formulas were

1) Common Intercept & Slope (Model I)

CGPA = ACTC

2) Separate Intercept & Common Slope (Model II)

CGPA = ACTC, HS-type

3) Separate Intercept & Slope (Model III)

CGPA = ACTC, HS-type, ACTC * HS-type

The same models as above were applied to the other two predictors (HSGPA and AA-HSGPA), replacing ACTC in the equations. Note also that each predictor was centered around its respective mean. This was done by subtracting a constant of 24.16, 3.41, or 2.57 from each student's ACTC score, HSGPA score, or AA-HSGPA score, respectively.

Results

Summary Statistics

Table 2 shows means and standard deviations for ACTC, HSGPA, AA-HSGPA and CGPA for Samples C, D, and E. Notice that easy and hard high schools are very similar in terms of ACTC, AA-HSGPA and CGPA, but different with respect to HSGPA (3.67 vs 3.16). Also note that AA-HSGPA is not on the same scale as HSGPA (i.e., not on a 0 to 4 scale).

Overall Difficulty of Courses

The overall difficulty of each course, obtained from Step 1 of the Scaling Analysis, is represented in Table 1 and plotted in Figure 3. Numbers on the overall curve in Figure 3 correspond to the high school course sequence number as shown in Table 1. Notice that

Computer Science (#12), English 11 (#3) and Beginning Calculus (#10) are ranked as easiest, moderate, and hardest overall, respectively.

School-specific Course Difficulty

Also, in Figure 3, the difficulty of the high school courses within two schools (denoted School A and School B) are plotted against overall average course difficulty. These schools were selected because they had the most extreme average course difficulty as measured by the courses' school-specific parameter estimates in Step 2 of the Scaling Analyses. School A's average course difficulty was –2.54 and school B's was 0.90. Overall average course difficulty was -0.22.

Step Difficulties for 50 high schools are depicted in Figure 4. Recall that the 23 high school courses were constrained to have a common set of step parameters within a school, but the step parameters were allowed to vary across schools. As shown in Figure 4, step thresholds get progressively higher as we move from step 1 to step 4.

Figure 5 shows the difficulty distribution of three selected courses across high schools. The selected courses were the easiest (Computer Science, #12) and hardest (Beginning Calculus, #10) overall according to Table 1, and also included a course of moderate overall difficulty (English 11, #3). The reliability of the school effect on these courses' difficulty varied from substantial (0.99 for English 11) to moderate (0.70 for Computer Science) to low (0.17 for Beginning Calculus). The reliability reflects differences in the across-school variance and in the percentage of students taking the course (the higher the percentage the higher the reliability for a given level of variance). The across-school variance was 1.14, .81, and .87 for, respectively, Computer Science, English 11, and Beginning Calculus. The percentage of students taking the course was, respectively, 8%, 86%, and 4% (see Table 1). Taken together, these results indicate

that grades are not generally comparable across high schools even for courses of the same title, and that grades in some courses, e.g., Computer Science, may be less comparable than grades in other courses, such as Beginning Calculus.

Midwest University Prediction Analyses

The proportion of variance (R^2) explained by each of the five prediction models was significant at the .01 level (see Table 3 for R^2 and p values). However, the AA-HSGPA was a better predictor than the HSGPA (0.25 versus 0.17). These R^2 values were improved (to 0.27 vs 0.22, respectively) by the addition of the ACT Composite as a predictor.

Dummy Variable Regression

Table 4 reports R^2 and the F-ratio for the difference between the three models I, II, and III regarding whether separate regression lines are needed for easy versus hard high schools. As expected there was a significant difference between hard and easy high schools regarding unadjusted high school GPA (HSGPA). This inference is based on the significant F-ratio for the difference between models II and I (5.49) and between models III and I (7.50). These results indicate that there is a significant main effect regarding high school type (easy or hard). However, in the presence of a significant F-ratio of 9.37 (at .01 level) for the difference between models III and II, we can conclude that separate intercept and slope is the appropriate model to be considered for predicting CGPA from HSGPA from high schools of different grading difficulty.

As expected there were no significant differences between easy and hard high schools in the regression coefficients for ACTC and AA-HSGPA. Neither main effects nor interaction effects were significant. This is apparent from Table 4, as none of the F-ratios for the difference between models were significant for the two predictors.

Once again the proportion of variance accounted for by each of the three predictors was significant at the .01 level. There was an increase in predictability measured by R^2 when the predictor variable was AA-HSGPA rather than raw high school GPA (HSGPA). This is apparent from Table 4 under R^2, when we compare the third and second rows.

Schematically the above inferences may be drawn from Figures 6, 7, and 8. As there is only significant high school type effect when predicting CGPA from HSGPA (see Figure 7) and no significant difference between easy and hard high schools when ACTC or AA-HSGPA are considered (see Figures 6 and 8, respectively). As expected this predicts that a student from a hard high school will have a higher CGPA than a student from an easy high school with an identical HSGPA. However, in the presence of interaction this difference becomes less and less apparent as we move to the high end of HSGPA. Note that the three points on each regression line corresponds to the mean, fifth percentile, and ninety-fifth percentile values of the respective predictor variable.

Conclusion

The results of this study reveal the problems with using the high school grade point average alone to predict achievement in college. As shown in Figure 7, HSGPA has a different relationship to college achievement, depending on the grading standards of the high school. This is the basic phenomenon that led early researchers to consider using high school and college-specific regressions in "central prediction systems" to improve information for college admissions decisions (Lindquist, 1963; Linn, 1966).

Our results also show and confirm the notion that a standardized test such as the ACT Assessment is a common yardstick for comparing academic achievement of students across schools. In Figure 6, the regression lines for easy vs. hard high schools are similar. In Table 4,

the parameters allowing separate intercept and slopes for easy versus hard high schools are not statistically significant.

The unique contribution of this study, however, is that it presents an additional way to use the ACT Assessment as a common yardstick. The traditional use of the ACT Assessment is to consider students' test scores, along with other factors, such as HSGPA, to predict students' achievement in college and to make college admissions decisions. This use concerns differences among students. Another way is to use the ACT Assessment to control for school effects (by constructing AA-HSGPA). This use concerns differences among schools. There is no overlap or redundancy between these two uses.

The continued, traditional use of ACT Assessment scores is consistent with the results of this study. Although grade-based, high school achievement measures, such as HSGPA and class rank (which is based on HSGPA) typically slightly outperform standardized tests such as the SATI or ACT in predicting college GPA (Noble 1991; Willingham, Lewis, Morgan, & Ramist, 1990; present study), such results have *not* been interpreted to suggest that college admissions decisions should be based on high school grades alone. One reason standardized tests continue to be used is that they *improve* the prediction of college GPA. In the present study, for example, R^2 values for predicting CGPA improved from 0.17 to 0.22 when ACTC was combined with HSGPA, and also improved from 0.25 to 0.27 when ACTC was combined with AA-HSGPA.

Another reason ACT Assessment tests should continue to be used traditionally is related to the nature of course grades. Course grades can be based on many factors such as attendance, social conformity, willingness to please authority figures, and class participation. These factors do not represent cognitive academic achievement. Indeed, one possible explanation for the fact that grade-based measures from high school and college correlate slightly better with each other

than with a standardized test is that they have more non-cognitive factors in common. Their widespread use as measures of academic achievement is sometimes attributed to the convenience of obtaining them rather than to their validity as measures of academic achievement (Johnson, 1997).

Finally, it seems to be common wisdom that no single variable can be relied upon for making admissions decisions or for representing academic achievement. Besides being comparable across schools, standardized tests are used in combination with grade data because they represent unique variables. As objective measures of cognitive academic achievement, ACT Assessment scores are less influenced by the same, possibly extraneous, factors that influence course grades. Therefore, even if high school grades could somehow be made comparable across schools, as with the AA-HSGPA, they would not represent the same variable as an ACT Assessment score. The consequence of this point is illustrated in this study by the fact that the ACTC improved the prediction of CGPA over that of AA-HSGPA alone.

The present study showed that scaling approaches to high school grade data could be useful. However, no adjustment was made on college grade data. Linn (1966) suggested using regression-adjusted GPA at both the high school and college levels in "central prediction systems." A related question for future research, then, is how much improvement in predicting achievement in college from high school grade data can be gained when adjustment is made at both levels. We expect that the best model will consist of AA-HSGPA and ACT Assessment scores predicting a scale-adjusted CGPA.

References

ACT (1997). *ACT Assessment technical manual.* Iowa City, IA: Author.

ACT (2002). *The high school profile report.* Iowa City, IA: Author.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 561-573.

Bashaw, W. L. (1965). Central-prediction-system models. *American Educational Research Journal, 2,* 151-162.

Bloom, B. S., & Peters, F. R. (1961). *Academic Prediction Scales.* New York: The Free Press of Glencoe.

Caulkins, J. P., Larkey, P. D., and Wei, J. (1996*). Adjusting GPA to Reflect Course Difficulty.* The Heinz School of Public Policy and Management, Carnegie Mellon University.

Johnson, V. E. (1997). An alternative to traditional GPA for evaluating student performance. *Statistical Science, 12,* 251-278.

Lei, P., Bassiri, D., & Schulz, E. M. (2003). *A Comparative Evaluation of Methods of Adjusting GPA in Differences in Grade Adjustment Practices. Journal of Applied Measurement,* (forthcoming).

Lindquist, E. F. (1963). An evaluation of a technique for scaling high school grades to improve prediction of college success. *Educational and Psychological Measurement, 23,* 623-646.

Linn, R. L. (1966). Grade adjustments for prediction of academic performance: A review. *Journal of Educational Measurement, 3,* 313-329.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149-174.

Maxey, J. (2001). *Selected trends on ACT-tested students.* Iowa City, IA: ACT.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*(2), 159-176.

Noble, J., & McNabb, T. (1989). *Differential coursework and grades in high school: Implications for performance on the ACT Assessment.* (ACT Research Report. 89-5). Iowa City, IA: ACT.

Noble, J. P. (1991). *Predicting college grades from ACT Assessment scores and high school course work and grade information.* (ACT Research Report. 91-3). Iowa City, IA: ACT.

Potthoff, R. F. (1964). *The prediction of college grades from College Board Scores and high school grades.* Institute of Statistics Mimeo Series, 419. Chapel Hill: University of North Carolina.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph,* 17.

Sawyer, R., Laing, J., & Houston, M. (1988). *Accuracy of self-reported high school courses and grades of college-bound students* (ACT Research Report. 88-1). Iowa City, IA: ACT.

Tucker, L. R. (1963). Formal models for a central prediction system. *Psychometric Monograph,10,* Richmond, VA: William Byrd Press.

Willingham, W. W., Lewis, C., Morgan, R., & Ramist, L. (1990). *Predicting college grades: an analysis of institutional trends over two decades.* Princeton, NJ: Educational Testing Service.

Wright, B. D., & Linacre, J. M. (1990). *A user's guide to BIGSTEPS: Rasch-model computer program version 2.0.* Chicago, IL: MESA Press.

Young, J. W. (1990). Adjusting the cumulative GPA using item response theory. *Journal of Educational Measurement, 27,* 175-186.

Young, J. W. (1993). Grade adjustment methods. *Review of Educational Research, 63*(2), 151-165.

Ziomek, R. L., and Svec, J. C. (1995). *High school grades and achievement: Evidence of grade inflation.* (ACT Research Report. 95-3). Iowa City, IA: ACT.

TABLE 1

List of First Twenty-Three Courses in the CGIS and Percent Course Taken and Course Difficulty from 50 High Schools

Course number	Course title	% taken	Rating scale model difficulty
1	English 9	94	-0.36
2	English 10	94	-0.39
3	English 11	86	-0.23
4	English 12	13	-0.77
5	Speech	35	-0.66
6	Algebra I	90	-0.18
7	Algebra II	79	0.16
8	Geometry	89	0.15
9	Trigonometry	33	0.29
10	Beginning Calculus	4	0.47
11	Other Advanced Math	16	0.27
12	Computer Science	8	-1.08
13	General Science	55	-0.60
14	Biology	92	-0.06
15	Chemistry	74	0.23
16	Physics	33	0.34
17	U.S. History	88	-0.31
18	World History	68	-0.21
19	Other History	19	-0.23
20	American Government	21	-0.39
21	Economics	19	-0.54
22	Geography	26	-0.65
23	Psychology	14	-0.42

TABLE 2

Descriptive Statistics for Models Predicting CGPA using ACTC, HSGPA and AA-HSGPA as Predictors by Different Samples of High Schools

	Sample			
Variable	D Students from easy high schools	E Students from hard high schools	D & E combined	C
ACTC	24.02 (3.44)	24.30 (3.09)	24.16 (3.27)	24.42 (3.52)
HSGPA	3.67 (0.31)	3.16 (0.41)	3.41 (0.36)	3.41 (0.46)
AA-HSGPA	2.68 (1.40)	2.47 (1.17)	2.57 (1.29)	2.70 (1.42)
CGPA	2.80 (0.66)	2.72 (0.61)	2.76 (0.64)	2.74 (0.67)

Note 1. SDs are given in parentheses. Also, SDs for combined samples are based on square root of pooled within variances of easy and hard high schools.
Note 2. AA-HSGPA is theoretically unbounded; ACTC is on 0 to 36 and HSGPA and CGPA both are on 0 to 4.

TABLE 3

R^2 Values for Models Predicting College GPA (CGPA)
N=1, 505 from 50 High Schools

Predictor	R^2	p
ACTC	0.12	0.0001
HSGPA	0.17	0.0001
ACTC, HSGPA	0.22	0.0001
AA-HSGPA	0.25	0.0001
ACTC, AA-HSGPA	0.27	0.0001

Note: The R^2 values in this table are somewhat lower than, or not strictly comparable to, those that are typically obtained or might be obtained. Typically, the R^2 between CGPA and a standardized test battery is based on multiple predictors (e.g., all four ACT tests), and the CGPA is the credit-weighted average of course grades. Likewise, in ACT's prediction research services, high school grades are represented as multiple predictors consisting of subject area GPA's in English, Mathematics, Social Studies, and Natural Sciences. In one recent ACT report (Maxey, 2001) the median correlation between the freshman CGPA and predictive indices based on either 1) ACT tests, 2) self-reported high school grades, or 3) their combination were, respectively, .423, .484, and .530. These translate to R^2 values of .18, .23, and .28, respectively. These values were based on 291 colleges and the 1997-98 college freshman cohort.

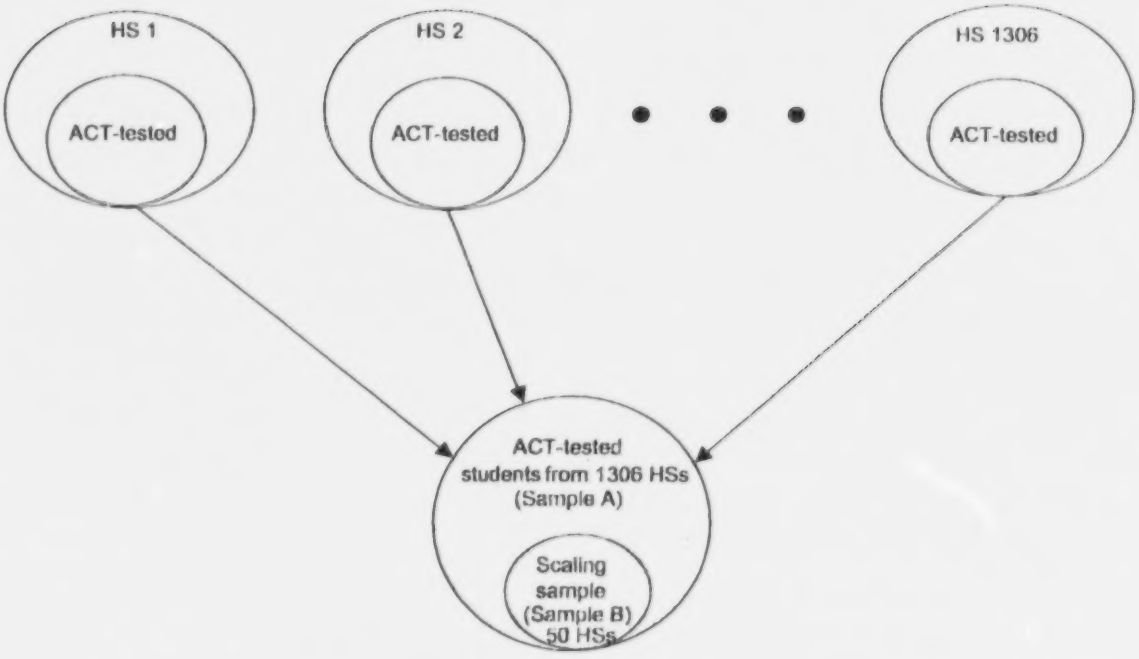
TABLE 4

Dummy Variable Regression for Easy and Hard High Schools

	R^2			F-ratio for difference		
Predictor	**Common intercept & slope (I)**	**Separate intercept & common slope (II)**	**Separate intercept & slope (III)**	**II vs I**	**III vs I**	**III vs II**
ACTC	0.10**	0.11**	0.11**	2.40	1.30	0.22
HSGPA	0.08**	0.10**	0.12**	5.49*	7.50**	9.37**
AA-HSGPA	0.16**	0.16**	0.16**	0.43	0.64	0.86

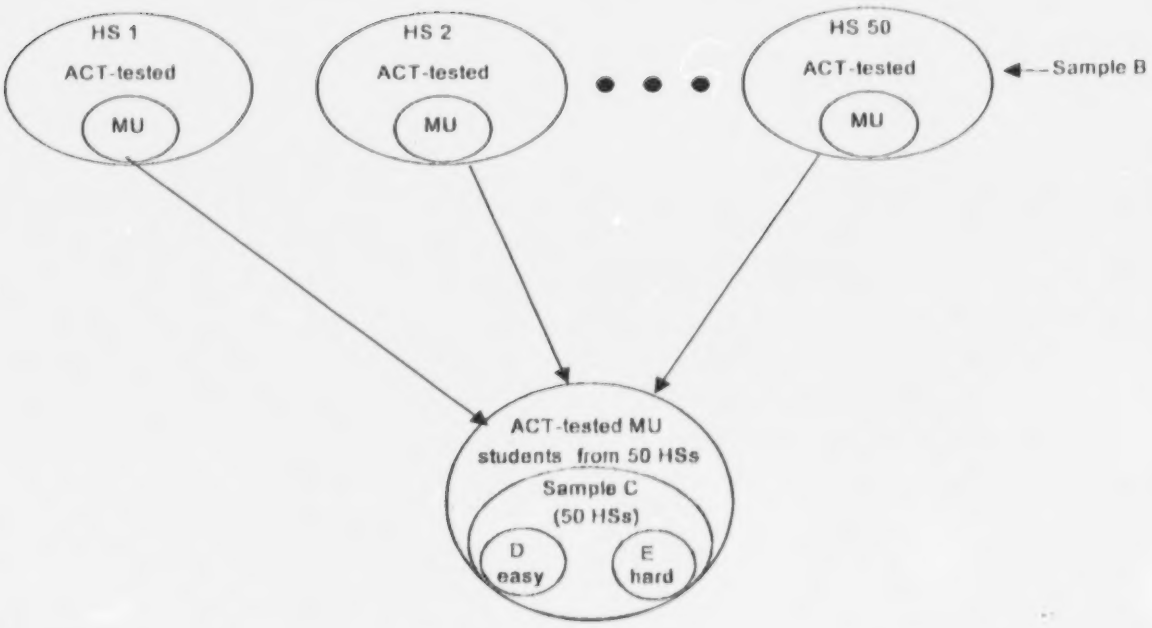
** Significant at .01 level
* Significant at .05 level

FIGURE 1. **ACT-Tested Samples (A & B)**



BEST COPY AVAILABLE

FIGURE 2. **ACT-Tested, Midwest University Samples (C, D, & E)**



BEST COPY AVAILABLE

FIGURE 3. **Course Difficulty Overall and Within Two High Schools with Extreme Average Course Difficulty**

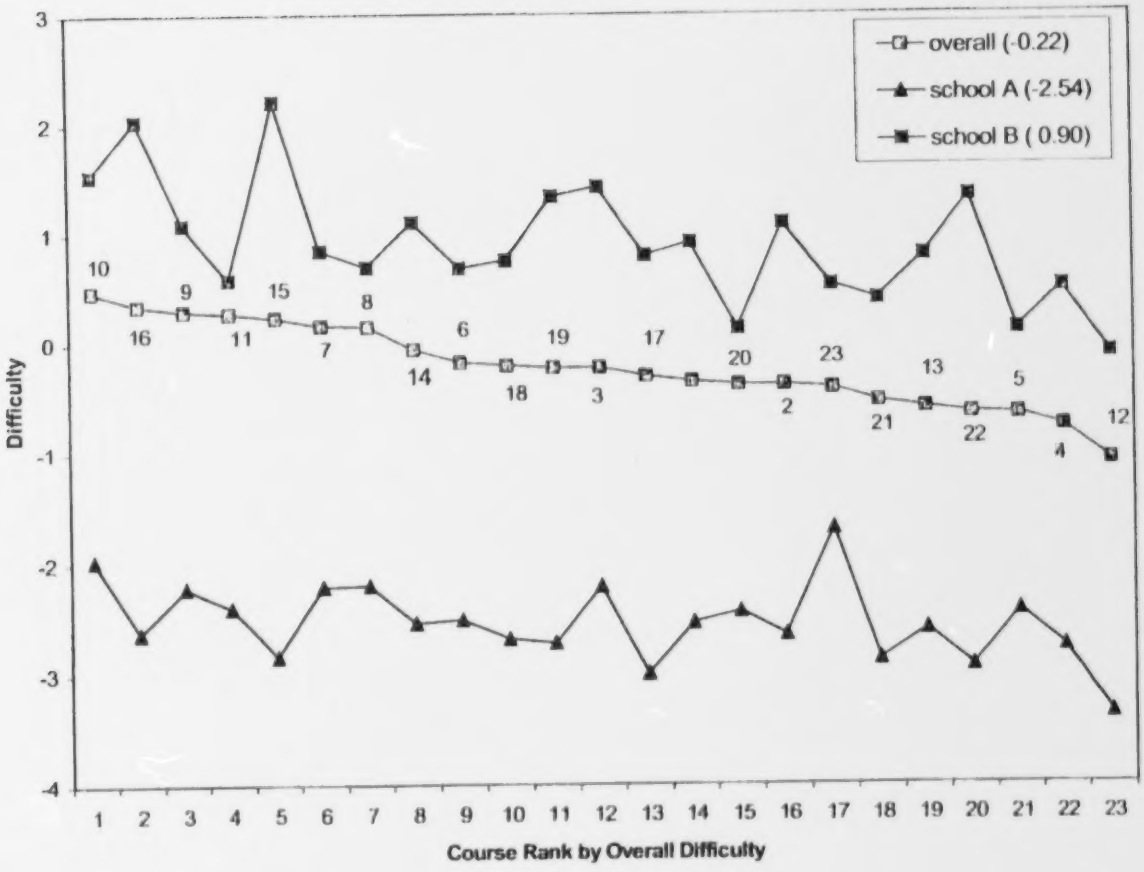


FIGURE 4. **Plot of Step Difficulties (τ_j) for 50 High Schools**

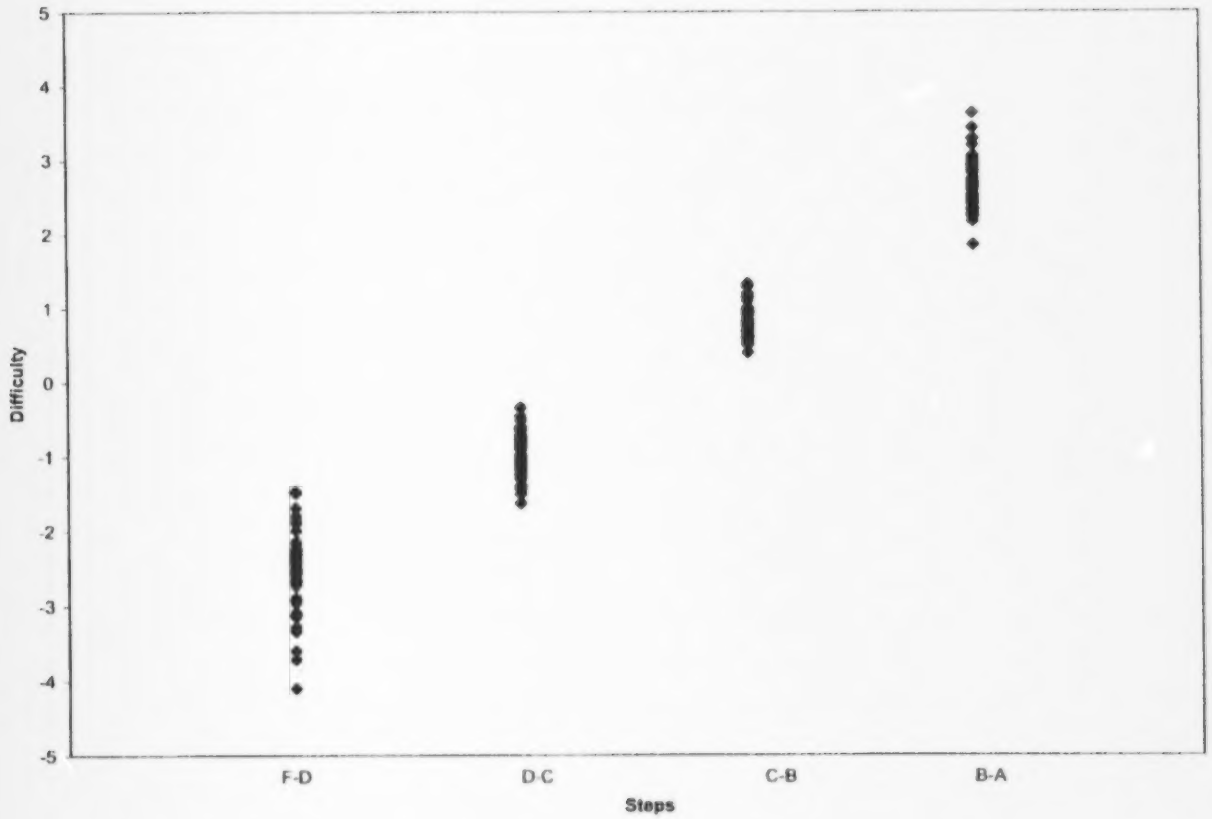


FIGURE 5. **Plot of Course Difficulties for 50 High Schools**

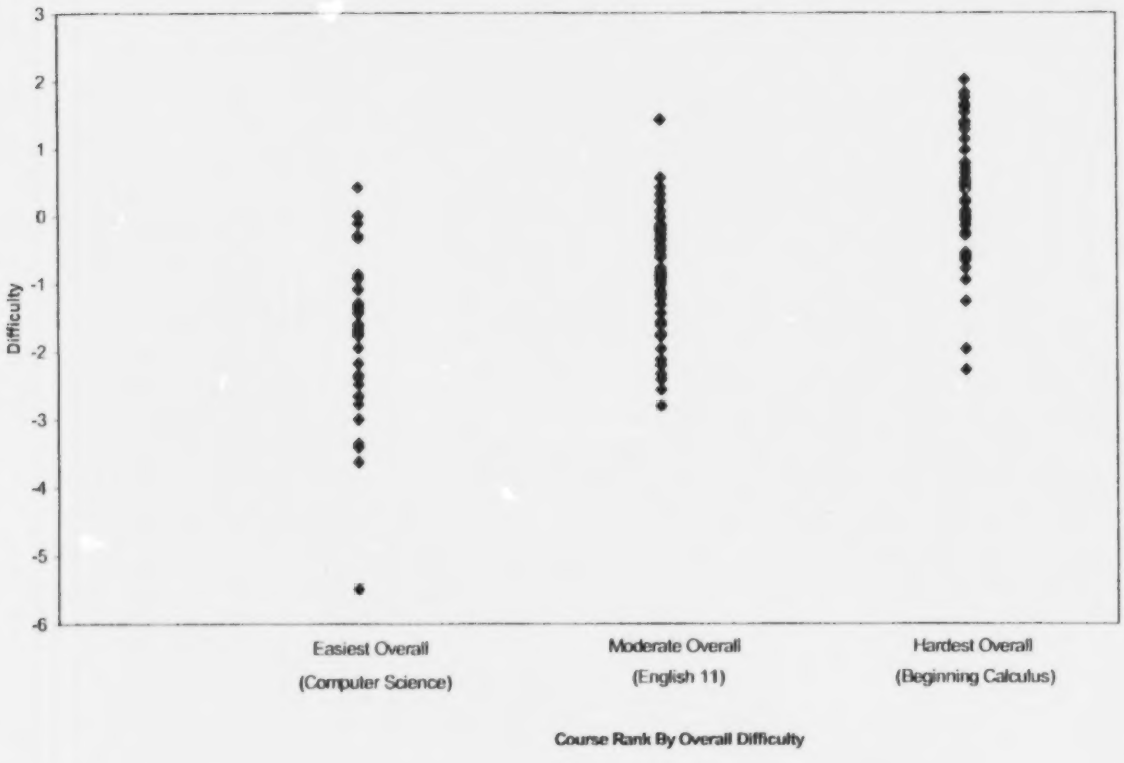


FIGURE 6. **Plot of Regression Line of CGPA on ACTC for Hard and Easy High Schools**

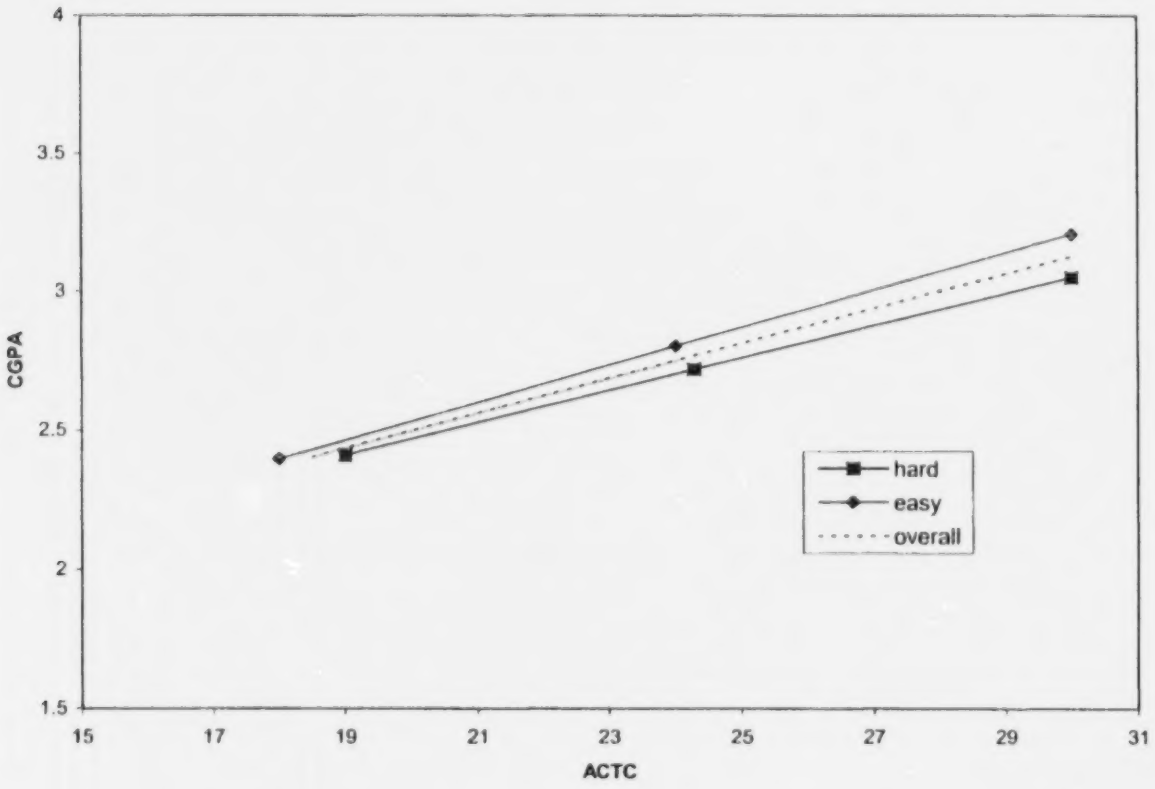


FIGURE 7. **Plot of Regression of CGPA on HSGPA for Hard and Easy High Schools**

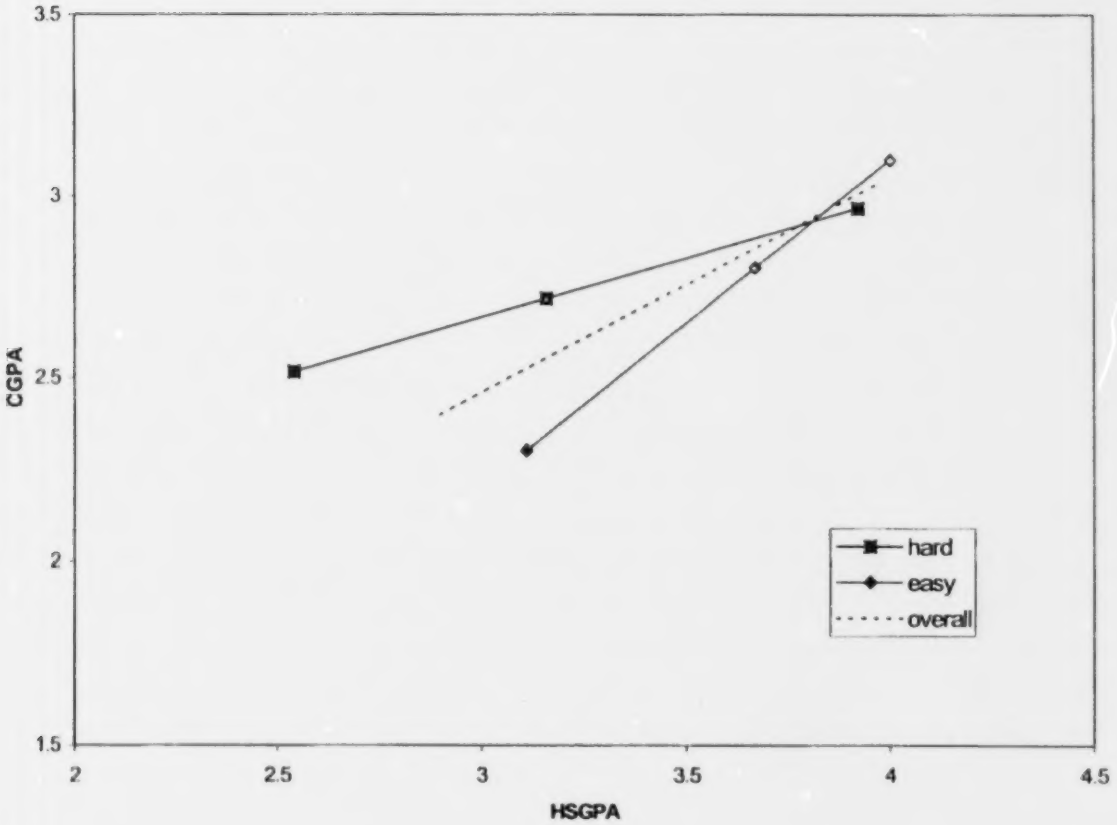
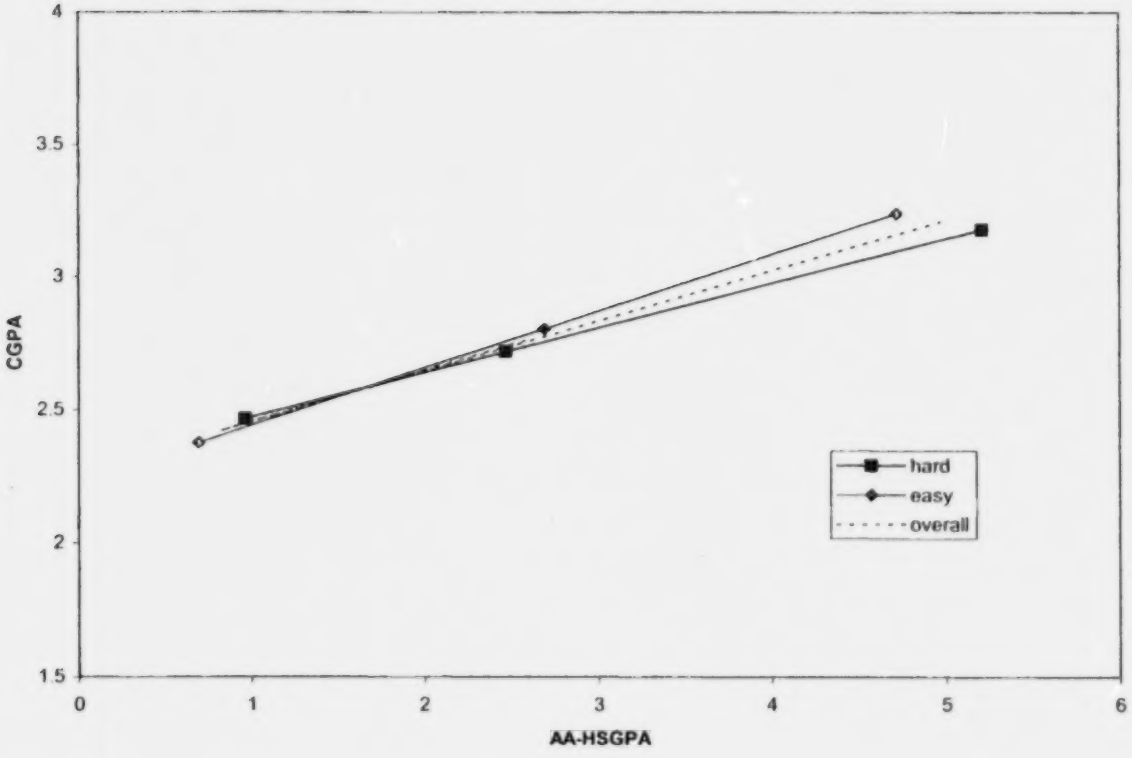


FIGURE 8. **Plot of Regression Line of CGPA on AA-HSGPA for Hard and Easy High Schools**



U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[X] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").